

A distributional perturbation method based on the Fisher-Rao distance for robustness analysis in uncertainty quantification

Baalu Belay Ketema^{†,1,2}, Nicolas Bousquet^{§,2}, Francesco Costantino^{§,1},
Fabrice Gamboa^{§,1}, Bertrand Iooss^{§,1,2}, Roman Sueur^{§,2}

[†] PhD student (presenting author). [§] PhD supervisors

PhD expected duration: Feb. 2023 – Feb. 2026

¹ Institut de mathématiques de Toulouse, Paul Sabatier University,
{baalu-belay.ketema, francesco.costantino, fabrice.gamboa}@univ-toulouse.fr

² EDF R&D,
{baalu-belay.ketema, roman.sueur, nicolas.bousquet, bertrand.iooss}@edf.fr

Abstract

Robustness analysis is a subdomain of uncertainty quantification that deals with the uncertainty propagation through a computer code G (assumed costly) of its input probability distributions. More specifically, the inputs (X_1, \dots, X_d) of G take uncertain values which usually correspond to physical measurements. Therefore, the uncertainty on these inputs X_i is modeled by a probability distribution f_i that is determined through physical experiments, hence f_i is itself uncertain. The goal of a robustness analysis method is then to: (a) take into account the uncertainty on f_i , this is done by defining a distributional perturbation method; (b) assess the impact on a quantity of interest (QoI) of the output $Y = G(\mathbf{X})$ of a perturbation on the inputs through robustness indices.

In this talk, we will present these two aspects of a specific robustness analysis method initially proposed in [1].

In the first part of this presentation, we will focus on the distributional perturbation method that is based on the Fisher-Rao distance on parametric families of probability distributions $\mathcal{P} = \{g_\theta\}_{\theta \in \Theta}$. This particular distance, on \mathcal{P} , derived from the Fisher information metric, has a geometric origin: it is the length of the shortest path connecting two points in \mathcal{P} . The Fisher-Rao distance presents many interesting properties for the purpose of robustness analysis. One of these properties is the universality of a (small) distance value $\delta > 0$ in two different families: if $\mathcal{P} = \{g_\theta\}_{\theta \in \Theta}$ and $\mathcal{Q} = \{h_\xi\}_{\xi \in \Xi}$ are two parametric families of probability distributions with their respective Fisher-Rao distance $d_{\mathcal{P}}$ and $d_{\mathcal{Q}}$, then if $d_{\mathcal{P}}(g_\theta, g_{\theta'}) = d_{\mathcal{Q}}(h_\xi, h_{\xi'}) = \delta > 0$ we can conclude that g_θ and $g_{\theta'}$ are “as far apart as” h_ξ and $h_{\xi'}$. This is a consequence of the Cramér-Rao lower bound and holds even though these distributions belong to two different families each possessing its own Fisher-Rao distance.

In the second part of this presentation, we will explain how the impact of a distributional perturbation on the inputs is assessed through robustness indices. These indices are denoted $S_{i\delta}$ for each input X_i and a perturbation level δ . They quantify the relative variation of the QoI of the output before and after perturbation of the input distributions

$$S_{i\delta} = \frac{Q(Y^{i\delta}) - Q(Y)}{Q(Y)},$$

where $Q(Y)$ is a QoI of Y , usually a quantile, and Y and $Y^{i\delta}$ are respectively the initial and perturbed output of G . Since these statistical quantities are not explicitly known, a statistical estimation method is necessary which takes into account the computation cost of G . The estimation method that we use is importance sampling. A central limit theorem is available for the estimators of both $Q(Y)$ and $Q(Y^{i\delta})$ but the asymptotic variance depends respectively on the density function of Y and $Y^{i\delta}$ which is hard to estimate [1, 2]. For this reason, we decided to construct non-asymptotic confidence intervals for the output QoI based on well known concentration inequalities [3]. This will be illustrated on different analytical and industrial cases.

Short biography (PhD student)

I did my bachelor's and master's degree in mathematics at Paul Sabatier University. I then pursued a PhD career in mathematics at EDF R&D in collaboration with the Mathematics Institute of Toulouse (IMT). The main goal in my thesis is to perform robustness analysis of computer codes for uncertainty quantification using information geometry. I am currently a second year PhD student and I was funded by the Labex CIMI (Toulouse) for the first few months of my PhD and by EDF for the remaining duration.

References

- [1] C. Gauchy, J. Stenger, R. Sueur, and B. Iooss. An information geometry approach to robustness analysis for the uncertainty quantification of computer codes. *Technometrics*, 64(1):80–91, 2022.
- [2] P. W. Glynn. Importance sampling for Monte Carlo estimation of quantiles. In *Mathematical Methods in Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation*, pages 180–185. Citeseer, 1996.
- [3] B. B. Ketema, N. Bousquet, F. Costantino, F. Gamboa, B. Iooss, and R. Sueur. Non-asymptotic confidence intervals for importance sampling estimators of quantiles. In *55ème Journées de statistique de la SFdS*, Bordeaux, France, May 2024.